

Digital humanities and libraries

This essay outlines a current trend in some academic libraries, specifically, the inclusion of digital humanities into their service offerings. It provides the briefest of introductions to the digital humanities, and then describes how one branch of the digital humanities -- text mining -- is being put into practice here in the Hesburgh Libraries' Center For Digital Scholarship at the University of Notre Dame.

Digital humanities

For all intents and purposes, the digital humanities is a newer rather than older scholarly endeavor. A priest named Father Busa is considered the "Father of the Digital Humanities" when, in 1965, he worked with IBM to evaluate the writings of Thomas Aquinas. With the advent of the Internet, ubiquitous desktop computing, an increased volume of digitized content, and sophisticated markup languages like TEI (the Text Encoding Initiative), the processes of digital humanities work has moved away from a fad towards a trend. While digital humanities work is sometimes called a discipline this author sees it more akin to a method. It is a process of doing "distant reading" to

evaluate human expression. (The phrase "distant reading" is attributed to Franco Moretti who coined it in a book entitled *Graphs, Maps, Trees: Abstract Models for a Literary History*. Distant reading is complementary to "close reading", and is used to denote the idea of observing many documents simultaneously.) The digital humanities community has grown significantly in the past ten or fifteen years complete with international academic conferences, graduate school programs, and scholarly publications.

Digital humanities work is a practice where digitized content of the humanist is quantitatively analyzed as if it were the content studied by a scientist. This sort of analysis can be done against any sort of human expression: written and spoken words, music, images,

dance, sculpture, etc. Invariably, the process begins with counting and tabulating. This leads to measurement, which in turn provides opportunities for comparison. From here patterns can be observed and anomalies perceived. Finally, predictions, thesis, and judgements can be articulated. Digital humanities work does not replace the more traditional ways of experiencing expressions of the human condition. Instead it supplements the experience.

This author often compares the methods of the digital humanist to the reading of a thermometer. Suppose you observe an outdoor thermometer and it reads 32° (Fahrenheit). This reading, in and of itself, carries little meaning. It is only a measurement. In order to make sense of the reading it is important to put it into context. What is the weather outside? What time of

year is it? What time of day is it? How does the reading compare to other readings? If you live in the Northern Hemisphere and the month is July, then the reading is probably an anomaly. On the other hand, if the month is January, then the reading is perfectly normal and not out of the ordinary. The processes of the digital humanist make it possible to make many measurements from a very large body of materials in order to evaluate things like texts, sounds, images, etc. It makes it possible to

evaluate the totality of Victorian literature, the use of color in paintings over time, or the rhythmic similarities & difference between various forms of music.

Digital humanities centers in libraries

As the more traditional services of academic libraries become more accessible via the Internet, libraries have found the need to necessarily evolve. One manifestation of this evolution is the establishment of digital humanities centers. Probably one of the oldest of these centers is located at the University of Virginia, but they now exist in many libraries across the country. These centers provide a myriad of services including combinations of digitization, markup,



Starry Night by Van Gogh

website creation, textual analysis, speaker series, etc. Sometimes these centers are akin to computing labs. Sometimes they are more like small but campus-wide departments staffed with scholars, researchers, and graduate students.

The Hesburgh Libraries' Center For Digital Scholarship at the University of Notre Dame was recently established in this vein. The Center supports services around geographic information systems (GIS), data management, statistical analysis of data, and text mining. It is located in a 5,000 square foot space on the Libraries's first floor and includes a myriad of computers, scanners, printers, a 3D printer, and collaborative work spaces. Below is an annotated list of projects the author has spent time against in regards to text mining and the Center. It is intended to give the reader a flavor of the types of work done in the Hesburgh Libraries:

- **Great Books** - This was almost a tongue-in-cheek investigation to calculate which book was the "greatest" from a set of books called the Great Books of the Western World.
- **HathiTrust Research Center** - The Center For Digital Scholarship facilitates workshops on the use of the HathiTrust Research Center as well as a small set of tools for programmatically searching and retrieving items from the HathiTrust.
- **JSTOR Tool** - JSTOR Tool is a Web-based application allowing the reader to summarize and do distant reading against data from JSTOR.
- **PDF To Text** - PDF To Text is a Web-based tool which extracts the plain text from PDF files and provides a number of services against the result.
- **Perceptions of China** - A faculty member leading the students on their trips to China wants to look for patterns of perception of trips to China
- **Poverty Tourism** - A university senior believes they have identified a trend -- the desire to tourist poverty-stricken places, and their investigations demonstrated the preponderant use of the word "you".
- **State Trials** - In collaboration with a number of other people, transcripts of the State Trials dating between 1650 and 1700 were analyzed.

- **Tiny Text Mining Tools** - A set of tiny text mining tools has been created enabling the Center to perform quick & dirty analysis against one or more items in a corpus.

Summary

Text mining, and digital humanities work in general, is simply the application computing techniques applied against the content of human expression. Their use is similar to use of the magnifying glass by Galileo. Instead of turning it down to count the number of fibers in a cloth (or to write an email message), it is being turned up to gaze at the stars (or to analyze the human condition). What he finds there is not so much truth as much as new ways to observe. The same is true of text mining and the digital humanities. They are additional ways to "see".

Eric Lease Morgan <emorgan@nd.edu>
University of Notre Dame
April 3, 2014

This essay and its companion blog posting (linked below) was written for the Information Organization Research Group, School of Information Studies at the University of Wisconsin Milwaukee



<http://blogs.nd.edu/emorgan/2014/04/dh-and-libraries/>